

The ethics of scholarly publishing: exploring differences in plagiarism and duplicate publication across nations*

Kathleen A. Amos, MLIS, AHIP

See end of article for author's affiliation.

DOI: <http://dx.doi.org/10.3163/1536-5050.102.2.005>

This study explored national differences in plagiarism and duplicate publication in retracted biomedical literature. The national affiliations of authors and reasons for retraction of papers accessible through PubMed that were published from 2008 to 2012 and subsequently retracted were determined in order to identify countries with the largest numbers and highest rates of retraction due to plagiarism and

duplicate publication. Authors from more than fifty countries retracted papers. While the United States retracted the most papers, China retracted the most papers for plagiarism and duplicate publication. Rates of plagiarism and duplicate publication were highest in Italy and Finland, respectively. Unethical publishing practices cut across nations.

INTRODUCTION

Published literature forms the basis of the scientific record; however, such literature is not without its flaws. Research errors occur, other researchers are unable to reproduce results, and scientific misconduct distorts the evidence on which future research is based. The retraction of publications provides one means of addressing such issues, correcting and ensuring the integrity of the literature [1]. While infrequent compared to the amount of literature published, data have shown recent increases in retractions in the biomedical literature [2–6], and numerous news articles and editorials have highlighted the need to investigate and retract papers [7–22]. The integrity of the published literature is of particular importance to librarians and other information professionals who endeavor to meet the information needs of health care providers, educators, researchers, patients, the public, and others. Librarians' success in fulfilling these needs depends on the available literature, and having reliable and accurate information is critical to research efforts, including systematic reviews; evidence-based practice; and decision making that can impact patient care and safety, the education of future health professionals, and the public's health.

Retraction is the most extreme method for addressing publication issues and, according to the Committee on Publication Ethics, "should usually be reserved for publications that are so seriously flawed (for whatever reason) that their findings or conclusions should not be relied upon" [1]. Among the reasons for taking this action are 2 issues that stem from unethical publishing behavior: plagiarism and duplication of findings that "have previously been published elsewhere without proper cross-referencing, permission or justification" [1]. Plagiarism has been estimated to account for 9.8%–17.0% of retractions, with duplicate

publication representing another 14.2%–17.0% [3, 5, 6, 23], and studies have suggested that retractions for plagiarism and duplicate publication have been increasing in recent years [3, 24].

Limited research has specifically considered whether there are differences between countries in retraction for plagiarism and duplicate publication in the biomedical literature cited in PubMed. Fang et al. reviewed articles retracted by May 2012 and found that authors from the United States, China, and India were responsible for the most plagiarism and duplicate publication retractions [3]. While comprehensive in terms of time period, the study was limited to English-language articles and did not provide the numbers of retractions recorded by countries for these two reasons. Additionally, differences in total numbers of retractions attributed to authors from various countries were not addressed, so it is not possible to determine whether a country's larger proportion of plagiarism or duplicate publication retractions was simply a reflection of a larger number of retractions. Stretton et al. analyzed plagiarism retractions through February 2008 and reported that, of authors with misconduct retractions, first authors from lower-income countries had higher odds of retraction for plagiarism than first authors from higher-income countries [24]. The research was again limited to English-language publications, among several other limitations, and did not distinguish plagiarism from self-plagiarism or duplicate publication.

This exploratory study focused on retractions in the biomedical literature due to plagiarism and duplicate publication, examining national differences in these practices. Specifically, two research questions were investigated:

- Which countries had the largest numbers of retractions for plagiarism and duplicate publication?
- Which countries had the highest rates of retraction for plagiarism and duplicate publication?

This exploration complements previous research by examining plagiarism and duplicate publication as distinct practices, estimating country-based rates of retraction for both of these practices in addition to numbers of retractions issued, and expanding the

* Based on a presentation at MLA '13, the 113th Annual Meeting of the Medical Library Association; Boston, MA; May 5, 2013.



Supplemental Table 1 is available with the online version of this journal.

study of retractions beyond literature published in English.

METHODS

To explore national differences in retraction for plagiarism and duplicate publication, the author performed an analysis of recent literature accessible through PubMed. A search of PubMed was conducted to identify all papers published from 2008 through 2012 that were later retracted, using the "Retracted Publication" publication type and a filter of "Publication date from 2008/01/01 to 2012/12/31." This search was run on January 27, 2013.

For each retrieved retracted paper, the author collected data on the national affiliations of the paper's authors, primarily from the paper itself or from the PubMed record if the information could not be obtained from the paper. Information was collected on all authors if available, and each paper was assigned a single country of authorship for the purposes of analysis. For papers with authors from multiple countries or with first authors affiliated with institutions in more than one country, the author based the analysis on the primary national affiliation of the first author.

To identify reasons for retraction, the author reviewed the retraction notices for the papers, and to a lesser extent the papers themselves. Papers were classified in one of four categories based on the reason for retraction: plagiarism, duplicate publication, other, or unknown. For the purposes of this analysis, plagiarism was defined as "the appropriation of another person's ideas, processes, results, or words without giving appropriate credit" [25], a definition used by the US Department of Health & Human Services' Office of Research Integrity. The definition of duplicate publication was derived from that of the National Library of Medicine: "an article that substantially duplicates another article without acknowledgement" when the "articles have one or more authors in common" [26]. Duplicative papers with no common authors that emerged from a single research group were also classified as duplicate publications. The category of "other" was used for the variety of other reasons for retraction, including error, inability to reproduce results, data fabrication or falsification, and lack of ethical approval for the research. "Unknown" was used when no reason for retraction was provided. Each paper was assigned to only one category, with papers coded as plagiarism or duplicate publication if either was indicated. Whenever possible, non-English-language retraction notices were translated using Google Translate.

Data on retracted publications were stored in Excel, and basic descriptive statistics were generated. Analysis for this study to determine numbers of retractions and retraction rates for plagiarism and duplicate publication was limited to countries with five or more retracted papers, or the equivalent of at least one paper per year on average, in order to focus on the countries contributing the most retractions to the

literature. Country-based retraction rates were determined by dividing the number of retractions for a particular reason by the total number of retractions issued for first authors from a given country.

RESULTS

The PubMed search for retracted papers published between 2008 and 2012 retrieved 835 papers, approximately 0.02% of the literature in PubMed for those years. Fourteen papers were excluded from further study due to incomplete information for analysis. The following findings are based on the remaining 821 papers. Within this sample, retractions were issued for authors from 53 countries, with the majority of countries responsible for very few retractions (Table 1, online only). Twenty countries had 5 or more retracted papers (Table 2), representing a total of 754 of the 821 retractions (91.8%). The United States had the most retracted papers at 199 (24.2% of all papers retracted), with China second at 143 (17.4%).

Of the 53 countries with retracted papers, 30 had retractions as a result of plagiarism and 34 as a result of duplicate publication (Table 1, online only). There was significant overlap in the countries facing these issues: 22 countries recorded both types of retractions. Plagiarism accounted for 136 of the 821 retractions (16.6%), while duplicate publication was implicated in 149 retractions (18.1%). The majority of papers (496, 60.4%) were retracted for reasons other than plagiarism or duplicate publication, and the reason for retraction could not be determined for 40 papers (4.9%).

Numbers of papers retracted due to plagiarism and duplicate publication

As shown in Table 2, of the 20 countries with the largest numbers of retractions, all had at least 1 paper retracted for either plagiarism or duplicate publication. Authors from China retracted the most papers due to both plagiarism and duplicate publication, withdrawing 24 plagiarized papers (17.6% of all papers retracted for plagiarism) and 42 duplicate papers (28.2% of all papers retracted for duplicate publication). With respect to plagiarism, India had the second highest number of papers retracted at 18, followed closely by the United States at 17 and Italy at 16.

In terms of duplicate publication, the United States had 26 retracted papers, second only to China. Japan and Germany followed, with 13 and 9 retracted papers, respectively.

Rates of retraction due to plagiarism and duplicate publication

Of the 20 countries that had retractions of 5 or more papers, the highest rate of retraction for plagiarism was found in Italy, where 66.7% of retractions resulted from plagiarism (Table 2). This was followed by Turkey at 61.5%, Iran and Tunisia at 42.9% each,

Table 2

Retraction numbers and rates for the 20 countries with 5 or more retracted papers, 2008–2012

Country	Number of papers retracted			Rate of retraction	
	Total	Plagiarism	Duplicate publication	Plagiarism	Duplicate publication
United States	199	17	26	8.5%	13.1%
China	143	24	42	16.8%	29.4%
Japan	57	2	13	3.5%	22.8%
Germany	55	0	9	—	16.4%
South Korea	55	3	7	5.5%	12.7%
India	49	18	7	36.7%	14.3%
United Kingdom	30	3	5	10.0%	16.7%
Italy	24	16	2	66.7%	8.3%
Australia	19	4	1	21.1%	5.3%
Netherlands	17	2	2	11.8%	11.8%
Iran	14	6	3	42.9%	21.4%
Canada	13	2	2	15.4%	15.4%
France	13	5	1	38.5%	7.7%
Turkey	13	8	2	61.5%	15.4%
Egypt	12	4	1	33.3%	8.3%
Spain	12	2	1	16.7%	8.3%
Brazil	9	3	1	33.3%	11.1%
Finland	8	0	3	—	37.5%
Tunisia	7	3	2	42.9%	28.6%
Sweden	5	1	0	20.0%	—

and France at 38.5%. In total, 12 countries had rates of plagiarism higher than the 16.6% average calculated for the sample. China's plagiarism rate was 16.8%, almost double the United States' rate of 8.5%. Both Finland and Germany recorded rates of 0.

For duplicate publication, fewer countries had retraction rates higher than the 18.1% sample average, and the range of rates was smaller. Finland had the highest rate of duplicate publication at 37.5%, followed by China at 29.4% and Tunisia at 28.6%. Japan (22.8%) and Iran (21.4%) also had rates above the sample average, while the rate of duplicate publication in the United States was below the average, at 13.1%. Only Sweden retracted no papers for duplicate publication.

DISCUSSION

Exploring plagiarism and duplicate publication across countries contributes to understanding publishing and retraction practices. Only a very small percentage of the published literature is ever retracted, and an even smaller percentage of that literature is retracted because of plagiarism or duplicate publication [3, 5, 6, 23]; however, these 2 reasons combined accounted for nearly 35% of all retractions in the studied sample. Duplicate publication appeared more common than plagiarism, with more papers retracted as a result of duplicate publication (149 versus 136) and more countries having duplicate publication retractions (34 versus 30). The average rates of retraction for plagiarism and duplicate publication in this study for the years 2008–2012, 16.6% and 18.1%, respectively, are comparable to those found in previous research [3, 5, 6, 23], but these aggregate numbers mask the fact that countries were not all affected by these ethical concerns to the same degree.

Just as countries varied widely in the number of retractions issued for their authors, from a high of 199 for the United States to a low of 1 for 15 countries,

variability was seen in the number of retractions stemming from plagiarism and duplicate publication among the 20 countries retracting the largest numbers of papers. China retracted the most literature for both of these reasons, although the United States contributed a significant number of plagiarism and duplicate publication retractions as well. Other countries retracting comparatively high numbers of papers for plagiarism included India and Italy and for duplicate publication included Japan, Germany, India, and South Korea, consistent with results reported by Fang et al. [3].

Investigating country-based rates of retraction for plagiarism and duplicate publication, however, offered a different perspective. While China, for example, retracted the most papers for both plagiarism and duplicate publication, its plagiarism retraction rate was only 16.8%. This was much lower than the plagiarism rates for Italy and Turkey, which both topped 60.0%. In terms of duplicate publication, China did not fare as well, with a rate of 29.4%, second only to Finland. Authors from the United States, on the other hand, may have had the most retracted literature, but plagiarism and duplicate publication did not seem to be the most critical concern. In both cases, the United States' rates were comparatively low, with only 8.5% of its retractions due to plagiarism and an additional 13.1% the result of duplicate publication. Retractions for other reasons far outweighed retractions for plagiarism and duplicate publication in this country.

Perhaps the most striking illustration of unethical publishing practices can be found by considering plagiarism and duplicate publication together. Although distinct, these practices stem from a similar root in that both are issues of originality. Whether authors duplicate someone else's work or their own, they are not contributing new material to the knowledgebase in the field. Combining rates of plagiarism and duplicate publication highlights even more

explicitly the originality problems occurring in select countries during this time period. Approximately 35% of the literature in the sample was retracted for plagiarism or duplicate publication, but more than 70% of the retracted literature by authors from Turkey, Italy, and Tunisia was duplicative in some way. Iran and India also had rates over 50%, and close to 50% of the retracted literature by Chinese authors was unoriginal. Several of these countries were also among those identified by Stretton et al. as having higher percentages of plagiarism retractions, using a definition that included self-plagiarism or duplicate publication [24]. The United States, while not free of originality issues, had a combined rate of plagiarism and duplicate publication that was only slightly over 20%.

Both retraction numbers and rates are important for assessing the extent of plagiarism and duplicate publication in countries. Because of differences in the amount of retracted literature available through PubMed, similar numbers can lead to widely varying rates of retraction and vice versa. For example, in this study, Tunisia's 3 retractions for plagiarism led to a plagiarism rate of 42.9%, while 3 retractions by authors from the United Kingdom translated to a rate of only 10.0%. Similarly, China's 16.8% plagiarism rate resulted from retraction of 24 papers, while Spain's rate of 16.7% resulted from only 2 retractions. By considering both numbers and rates, a broader view of the problem emerges. Although plagiarism and duplicate publication may not be the cause of most retractions for US authors, they are not uncommon in terms of numbers.

LIMITATIONS OF THE STUDY

Several limitations should also be taken into account in considering these findings. First, this study provides a snapshot of retracted biomedical literature over a single time period, 2008–2012, that was available through a single database, PubMed. Further research would be needed to determine whether the same results are seen for all retracted papers in PubMed or for papers available through other databases. In addition, it was not always possible to access all of the information on a paper needed for this study, although less than 2% of the papers retrieved in the PubMed search were excluded from analysis.

The study also did not consider differences in the amount of literature published by authors from various countries, only the amount of literature retracted. Preliminary research suggests that examining plagiarism and duplicate publication as a percentage of the published literature offers yet another view on the national distribution of these practices. Further comparative research is needed to explore these differences.

In conducting this analysis, the author coded papers by the primary national affiliation of the first author; however, at least 18.8% (154) of the papers in this sample were written by authors from 2 or more

countries. Considering the impact of multinational authorship could provide a different perspective, as could considering the impact of individual authors. Despite the sometimes high rates of plagiarism and duplicate publication found, numbers of retractions for all countries were small and might have been influenced by the retraction of multiple papers by the same author. Exploration into whether plagiarism and duplicate publication are widely distributed or confined to a small number of authors is warranted.

Finally, papers were assigned only one reason for retraction, retraction notices were not always clear, and assumptions were avoided. Author names and institutions were used to distinguish plagiarism from duplicate publication, but it is possible that two authors with the same name could have led to a classification of duplicate publication when a classification of plagiarism would have been more appropriate. In addition, where plagiarism or duplicate publication were not clearly apparent from the retraction notices, papers were not classified as such. The results likely underestimate the true rates of plagiarism and duplicate publication in retractions.

CONCLUSION

This exploratory study adds to the literature on retractions by including non-English-language papers, differentiating plagiarism from duplicate publication, and estimating country-based rates of retraction for plagiarism and duplicate publication in addition to numbers of retractions issued for these reasons. Although similar, plagiarism and duplicate publication may be perceived differently in terms of level of severity, and retraction numbers and rates provide different perspectives on the extent of plagiarism and duplicate publication in countries. No country is unique in having to address issues of plagiarism and duplicate publication, although such unethical behaviors may be a more pressing concern for some countries than others. This may suggest the need for different educational strategies related to publishing ethics or other means of ensuring publishing integrity in different countries. Continuing research with a larger sample; considering differences among countries in the amount of literature published; and investigating the numbers of authors involved, and not only countries, would shed further light on the scope of the problem.

REFERENCES

1. Wager E, Barbour V, Yentis S, Kleinert S, COPE Council. Retractions: guidance from the Committee on Publication Ethics (COPE). *Obes Rev*. 2010 Jan;11(1):64–6. DOI: <http://dx.doi.org/10.1111/j.1467-789X.2009.00702.x>.
2. Cokol M, Ozbay F, Rodriguez-Esteban R. Retraction rates are on the rise. *EMBO Rep*. 2008 Jan;9(1):2. DOI: <http://dx.doi.org/10.1038/sj.embor.7401143>.
3. Fang FC, Steen RG, Casadevall A. Misconduct accounts for the majority of retracted scientific publications. *Proc Natl Acad Sci U S A*. 2012 Oct 16;109(42):17028–33. DOI: <http://dx.doi.org/10.1073/pnas.1212247109>.

4. National Library of Medicine. Key MEDLINE indicators [Internet]. Bethesda, MD: The Library; 7 Feb 2002 [rev. 30 Nov 2012; cited 13 Oct 2013]. <http://www.nlm.nih.gov/bsd/bsd_key.html>.
5. Steen RG. Retractions in the scientific literature: is the incidence of research fraud increasing? *J Med Ethics*. 2011 Apr;37(4):249–53. DOI: <http://dx.doi.org/10.1136/jme.2010.040923>.
6. Wager E, Williams P. Why and how do journals retract articles? an analysis of Medline retractions 1988–2008. *J Med Ethics*. 2011 Sep;37(9):567–70. DOI: <http://dx.doi.org/10.1136/jme.2010.040964>.
7. A retrospective of retractions: the striking record in 2011. *Nat Med*. 2011 Dec 6;17(12):1544. DOI: <http://dx.doi.org/10.1038/nm1211-1544>.
8. Balter M. Reviewer's déjà vu, French science sleuthing uncover plagiarized papers. *Science*. 2012 Mar 9;335(6073):1157–8. DOI: <http://dx.doi.org/10.1126/science.335.6073.1157>.
9. Bartold PM. Plagiarism—it's just not on. *Aust Dent J*. 2011 Dec;56(4):347. DOI: <http://dx.doi.org/10.1111/j.1834-7819.2011.01373.x>.
10. Collier R. Shedding light on retractions. *CMAJ*. 2011 Apr 19;183(7):E385–6. DOI: <http://dx.doi.org/10.1503/cmaj.109-3827>.
11. Copy and paste [editorial]. *Nature*. 2011 May 26;473(7348):419–20. DOI: <http://dx.doi.org/10.1038/473419b>.
12. Couzin-Frankel J. Shaking up science. *Science*. 2013 Jan 25;339(6118):386–9. DOI: <http://dx.doi.org/10.1126/science.339.6118.386>.
13. Cyranoski D. Retraction record rocks community. *Nature*. 2012 Sep 20;489(7416):346–7. DOI: <http://dx.doi.org/10.1038/489346a>.
14. Cyranoski D. Zero tolerance. *Nature*. 2012 Jan 12;481(7380):134–6. DOI: <http://dx.doi.org/10.1038/481134a>.
15. Dyer C. Japanese doctor is heading for record number of retracted research papers. *BMJ*. 2012 Jul 4;345:e4596. DOI: <http://dx.doi.org/10.1136/bmj.e4596>.
16. Jones DS. Plagiarism and misconduct in scientific publishing. *J Pharm Pharmacol*. 2009 May;61(5):539. DOI: <http://dx.doi.org/10.1211/jpp/61.05.0001>.
17. Poster E, Pearson GS, Pierson C. Publication ethics: its importance to readers, authors, and the profession. *J Child Adolesc Psychiatr Nurs*. 2012 Feb;25(1):1–2. DOI: <http://dx.doi.org/10.1111/j.1744-6171.2011.00317.x>.
18. Reich ES. Symmetry study deemed a fraud. *Nature*. 2013 May 9;497(7448):170–1. DOI: <http://dx.doi.org/10.1038/497170a>.
19. Tramèr MR. The Boldt debacle. *Eur J Anaesthesiol*. 2011 Jun;28(6):393–5. DOI: <http://dx.doi.org/10.1097/EJA.0b013e328347bdd1>.
20. Van Noorden R. Science publishing: the trouble with retractions. *Nature*. 2011 Oct 6;478(7367):26–8. DOI: <http://dx.doi.org/10.1038/478026a>.
21. Van Steirteghem A, Williams AC. Plagiarism, retraction and the future. *Hum Reprod*. 2011 Aug;26(8):1940. DOI: <http://dx.doi.org/10.1093/humrep/der219>.
22. Yentis SM. Another kind of ethics: from corrections to retractions. *Anaesthesia*. 2010 Dec;65(12):1163–6. DOI: <http://dx.doi.org/10.1111/j.1365-2044.2010.06557.x>.
23. Redman BK, Yarandi HN, Merz JF. Empirical developments in retraction. *J Med Ethics*. 2008 Nov;34(11):807–9. DOI: <http://dx.doi.org/10.1136/jme.2007.023069>.
24. Stretton S, Bramich NJ, Keys JR, Monk JA, Ely JA, Haley C, Woolley M, Woolley KL. Publication misconduct and plagiarism retractions: a systematic, retrospective study. *Curr Med Res Opin*. 2012 Oct;28(10):1575–83. DOI: <http://dx.doi.org/10.1185/03007995.2012.728131>.
25. Office of Research Integrity, US Department of Health & Human Services. Definition of research misconduct [Internet]. Rockville, MD: The Office; [updated 25 Apr 2011; cited 13 Oct 2013]. <<http://www.ori.dhhs.gov/definition-misconduct/>>.
26. National Library of Medicine. Fact sheet: errata, retractions, partial retractions, corrected and republished articles, duplicate publications, comments (including author replies), updates, patient summaries, and republished (reprinted) articles policy for MEDLINE [Internet]. Bethesda, MD: The Library; 6 Mar 2002 [rev. 27 Nov 2012; cited 9 Oct 2013]. <<http://www.nlm.nih.gov/pubs/factsheets/errata.html>>.

AUTHOR'S AFFILIATION

Kathleen A. Amos, MLIS, AHIP, kamos@phf.org, Project Manager, Council on Linkages Between Academia and Public Health Practice, Public Health Foundation, 1300 L Street Northwest, Suite 800, Washington, DC 20005

Received October 2013; accepted November 2013